SINGLE WHEAT KERNEL COLOR CLASSIFICATION USING NEURAL NETWORKS

D. Wang, F. E. Dowell, R. E. Lacey

ABSTRACT. *An optical radiation measurement system, which measures reflectance spectra from 400 to 2000 nm, was used to quantify single wheat kernel color. Six classes of wheat were used for this study. A neural network (NN) using input data dimension reduction by divergence feature selection and by principal component analysis was used to determine single wheat kernel color class. The highest classification accuracy was 98.8% when the divergence feature selection method was used to reduce the number of NN inputs. The highest classification accuracy was 98% when principal component analysis method was used to reduce the number of NN inputs.*

Keywords. *Wheat, Color, Principal component analysis, Artificial neural networks, Near-infrared reflectance.*

Wheat is classified as either red or white depending on the color of the seed coat. These two basic colors, as well as certain variations within each color, are commonly considered in the classification of wheat for grading purposes. Red wheat and white wheat have different milling, baking, and taste properties, and different visual characteristics. In the world markets, a premium may be paid for a particular color class based on nutritional and end-use values (Ronalds and Blakeney, 1995; Bason et al., 1995). Currently, wheat color is determined by licensed inspectors. In the United States, USDA Grain Inspection Packers and Stockyard Administration (GIPSA) personnel visually examine wheat samples to determine kernel color. Inspectors may use slightly different criteria to distinguish red from white wheats. This can result in unreliable classification when the threshold between red and white is not clear. Wheat color is controlled by up to three red genes (Metzger and Sibaugh, 1970). Baker (1981) indicated that various combinations of those red genes impart different shades of red to the genotypes. Flintham (1993) reported that the degree of red pigmentation increases with the number of the genes (one to three). This would make some single red gene wheats difficult to distinguish from white wheat. Beside genetics, weather conditions such as rain damage which is known to "bleach" red wheat, soil conditions, cultivar, disease, and insect damage frequently cause variations within each color class and affect visual appearance of the kernel. Therefore, red and white wheats are not always visually distinguishable. Thus, wheat from each color class may get mixed, resulting in wheat of lower quality and value than the pure lots.

Several methods to measure wheat color have been studied. Kernels can be soaked in a solution of sodium hydroxide (NaOH) to assist inspectors in determining color class. Genetically red kernels tend to turn red after soaking; whereas, genetically white kernels tend to turn light cream in color (Dowell, 1997; Chmelar and Mostovoj, 1938; Quartley and Wellington, 1962; Kimber, 1971; DePauw and McCaig, 1988). Chen et al. (1972) converted perceived color to a point in a three-dimensional color space using a colorimeter. Other methods quantify kernel color by measuring reflectance at many different wavelengths (Massie and Norris, 1965; Hawk et al., 1970; McCaig et al., 1992, 1993; Ronalds and Blakeney, 1995; Delwiche and Massie, 1996). However, most research measured the color of bulk samples. Quantifying the color of single kernels is necessary to determine whether a bulk sample has a mixture of red and white wheat classes. Moreover, most previous research did not include kernels that were not obviously red or white. Thus, some means of measuring color of single kernels, including kernels that are not obviously red or white, is needed.

Near-infrared reflectance may provide a means of rapidly classifying genetically red and white wheats, independent of visual characteristics. The acceptance of near-infrared (NIR) reflectance analysis as a quantitative tool in recent years is, in large part, the result of increases in the quality of the spectra and the power of the mathematics used in analysis. For quantitative analysis, the mathematical techniques of multiterm linear regression (MLR) (Hruschka, 1990), partial least squares (PLS) regression (Bjørsvik and Martens, 1992), and principal component regression (PCR) (Gunst and Mason, 1979) have been used for extracting the necessary information from the resulting mass of data. Artificial neural networks (ANN) can also be used for quantitative analysis. Dowell (1994) used ANN to classify damaged and undamaged peanuts kernels. The results showed that, when compared

Article was submitted for publication in March 1998; reviewed and approved for publication by the Information & Electrical Technologies Division of ASAE in November 1998.

Mention of trademark or proprietary product does not constitute a guarantee or warranty of the product by the U.S. Dept of Agric. and does not imply its approval to the exclusion of other products that also may be suitable.

The authors are **Donghai Wang,** Research Associate, and **Floyd E. Dowell,** ***ASAE Member Engineer,*** Agricultural Engineer, USDA-ARS, Grain Marketing and Production Research Center, Manhattan, Kans.; and **Ronald E. Lacey,** ***ASAE Member Engineer,*** Associate Professor, Agricultural Engineering Dept., Texas A&M University, College Station, Tex. **Corresponding author:** Floyd E. Dowell, USDA-ARS, Grain Marketing and Production Research Center, 1515 College Ave., Manhattan, KS 66502; voice: (785) 776-2753; fax: (785) 776-2792; e-mail: fdowell@usgmrl.ksu.edu.

to statistical means of classifying kernels using specific wavelengths or data from a colorimeter, ANN correctly classified about 5% and 13% more kernels for damaged and undamaged kernels, respectively. Delwiche and Norris (1993) used PCR to classify ground samples of hard red spring and hard red winter with a correct classification rate of 95%. Using the same data, Chen et al. (1995) developed a back-propagation neural network classification model with correct classification rate of 95.9%. Song et al. (1995) used neural network models to determine wheat class with an average accuracy of 98% to 100% for two-class models and 95% for a six-class models. However, for color classification, most results were based on PLS and MLR analysis of spectra from bulk samples. For neural networks, the large data dimension (number of wavelengths used for analysis) is a major problem which limits the application of neural network to spectral data analysis. The large number of input nodes requires many training examples and excessive training time for developing a model with desired prediction capability.

The objectives of this research were to: (1) investigate the feasibility of wheat color classification based on NIR reflectance spectra of single kernels using a neural network technique; and (2) find an effective method to reduce the spectra data dimension for neural network analysis.

Materials and Methods

Materials

Six U.S. market classes of wheat, hard red spring (HRS), hard red winter (HRW), soft red winter (SRW), hard white wheat (HWW), soft white wheat (SWW), and durum (DUR) supplied by the GIPSA technical center (Kansas city, Mo.), were used for single wheat kernel color classification (table 1). Each class was represented by six or seven varieties. Twenty-five kernels were randomly selected from each variety for a total of 150 or 175 kernels per class. In addition, 200 kernels were selected from wheats determined to be difficult-to-classify (DTC) as red or white by USDA GIPSA. Most durum wheat in the U.S. is genetically white. Red durum wheat is grown to a very limited extend for feed purpose only. Therefore, only white durum wheats were used in this research. Samples originated from the 1993-1995 crop years.

Equipment and Operation

An optical radiation measurement system (Oriel Corporation, Stratford, Conn.) was used to measure single wheat kernel reflectance spectra from 400 to 2000 nm at 2 nm intervals. The optical radiation measurement system consisted of a radiation source system, a control system (RS-232 Merlin™, Model No. 70100), a chopper (Model No. 75152), a monochromator (Model No. 77250 1/8 m), a lead sulfide detector (Model No. 70131), and software (fig. 1). The radiation source system consisted of a 12 V power supply (Model No. 68831), a photofeedback controller system (Model No. 68550), a 100 W lamp (Model No. 6333), and lamp housing (Model No. 60000). The power supply was designed to meet the need for a highly regulated source of current for proper operation of the quartz tungsten halogen lamp. The photofeedback controller system interfaced with the power supply to maintain a constant radiation level regardless of lamp aging, line voltage variation, and change in ambient temperature. The control system controlled the chopper speed and received signals from the detector. This unit was designed to provide all the necessary functions needed for the phase sensitive detection of low radiation signals and utilized state of the art digital processing techniques. The chopper modulated the radiation to be measured. The monochromator provided a wide spectral range (190 to 2400 nm) with interchangeable gratings, high throughput, and low stray radiation. The software integrated all individual parts in the system and controlled the individual components from a PC-compatible computer running the Windows operating system.

Chopped light was delivered by a 2-mm diameter optical liquid fiber (Model No. 77634) with an F# = 1 to illuminate a single wheat kernel. The illumination fiber was oriented 45° from vertical, with the fiber end being located 8-mm from the kernel. Another 2-mm diameter optical liquid fiber (Model No. 77634) with an F# = 1 collected radiation reflected from wheat kernel. The reflectance fiber was oriented vertically above the wheat kernel with the fiber end being located 6 mm from the wheat kernel. The grating (Model No. 77350) was blazed at a wavelength of 750 nm with a groove density of 600 lines/mm and a reciprocal dispersion of 13.2 nm/mm. The primary grating wavelengths were 450 to 2000 nm with a usable range of

Table 1. Wheat samples used for single wheat kernel color classification

Class*	Varieties	n†	MC (%)†	Protein (%)†	Hardness†
HRS	6	150	10.70-12.50	10.11-17.30	63.6-90.3
HRW	6	150	9.42-11.72	9.38-16.16	63.6-81.2
SRW	6	150	10.81-15.26	8.47-11.50	13.9-25.3
HWW	6	150	9.07-12.07	10.63-14.03	59.9-72.5
SWW	7	175	9.32-15.93	8.70-14.24	22.6-49.8
DUR	6	150	11.43-12.46	12.30-16.90	87.0-94.8
DTC	8	200			

* HRS = Hard red spring, HRW = Hard red winter, SRW = Soft red winter, HWW = Hard white wheat, SWW = Soft white wheat, DUR = Durum, and DTC = Difficult to classify.

† n = number of kernels per class (25 kernels per variety). MC, protein, and hardness were average value of the bulk samples.

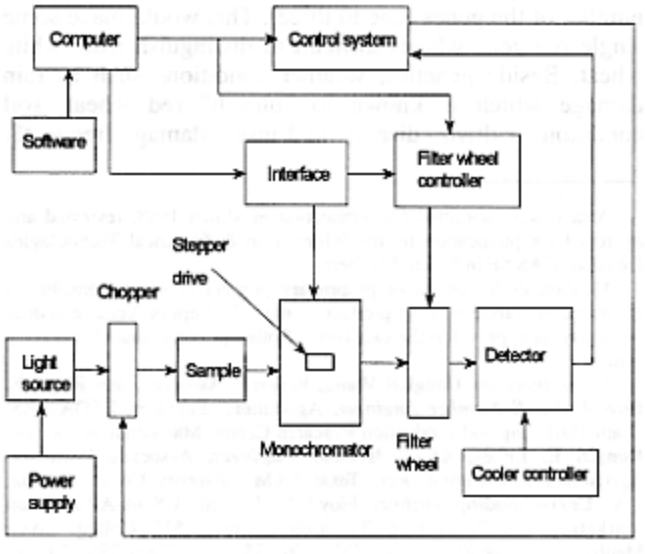

Figure 1–An optical radiation measurement system used to measure single wheat kernel reflectance.

400 to 2000 nm. Identical entrance and exit slits for the monochromator were set at a height of 12 mm and a width of 1.56 mm. This gives a passband of 10 nm. A lead sulfide detector measured the single wheat kernel reflectance from 400 to 2000 nm. Reflectance readings were referenced to a standard baseline created using spectralon™ with 99% reflectance (Labsphere Inc., North Sutton, N.H.). To minimize instrument fluctuations over time, the standard baseline was scanned once per hour. The constant time and pause time of the detector were set at 0.3 and 1.0 s, respectively.

Data Pretreatments

The spectral data were first transformed to Log(1/R) using the software package "Gram/32" (Galactic Industries Co., Salem, N.H.) and stored in Grams/32 format. The spectral data were smoothed using the Savitsky-Golay smoothing method (Savitsky and Golay, 1964). This method uses a convolution approach which performs a least squares fit to a specified window of data points. Smoothing is controlled by the degree of polynomial and the number of smoothing points parameters. The degree of polynomial specifies the order of the polynomial to fit over the specified number of smoothing points. A 5th-order polynomial with 25 points was used for smoothing the wavelength region from 400 to 1450 nm. Because more noise occurred after 1450 nm wavelength, a fifth-order polynomial with 45 points was used for smoothing the wavelength region from 1450 to 2000 nm. The calculation of the first and second derivatives was also based on the Savitsky-Golay method. Derivatives are an approach to addressing two basic problems with near-infrared spectra: overlapping peaks and large baseline variations (William, 1990). A fifth-order polynomial with 25 points was used to calculate the first and second derivatives. Mean centering was used to scale the data so that the mathematics of the spectral decomposition and correlations would perform better. This involves calculating the average of all the spectra in the training set and then subtracting the average from each spectrum. By removing the mean from the data, the differences between the samples are substantially enhanced in terms of both concentration and spectral response.

Divergence Feature Selection

Divergence is a measure of "distance" or dissimilarity between two classes. It can be used to determine feature ranking and to evaluate the effectiveness of class discrimination. If there are two classes ω_1 and ω_2 and their probability density functions at X are $P(X/\omega_1)$ and $P(X/\omega_2)$, then the discriminating information for class ω_1 and ω_2 at X is defined as:

$$u_{12} = \ln \frac{P\left(\frac{X}{\omega_1}\right)}{P\left(\frac{X}{\omega_2}\right)} = -u_{21} \qquad (1)$$

The average discriminating information is:

$$I_{(1,2)} = \int P\left(\frac{X}{\omega_1}\right) \ln \frac{P\left(\frac{X}{\omega_1}\right)}{P\left(\frac{X}{\omega_2}\right)} dx \qquad (2)$$

The inverse average discriminating information is:

$$I_{(2,1)} = \int P\left(\frac{X}{\omega_2}\right) \ln \frac{P\left(\frac{X}{\omega_2}\right)}{P\left(\frac{X}{\omega_1}\right)} dx \qquad (3)$$

The divergence is the sum of the discriminating information:

$$J_{(1,2)} = I_{(1,2)} + I_{(2,1)} \qquad (4)$$

Assume that both classes are described by normal distributions:

$$N(m_1, C_1) \quad \text{and} \quad N(m_2, C_2)$$

where m_1 and m_2 are the mean vectors, and C_1 and C_2 are n by n covariance matrices. The divergence for two classes is:

$$J_{(1,2)} = \frac{1}{2} \mathrm{tr}\left[(C_1 - C_2)(C_2^{-1} - C_1^{-1}) \right]$$

$$= \frac{1}{2} \mathrm{tr}\left[(C_1^{-1} + C_2^{-1})(m_1 - m_2)(m_1 - m_2)' \right] \qquad (5)$$

where tr is the trace (the sum of the diagonal elements of matrix). Two special cases are of particular interest:

Case 1. Equal covariance matrices: $C_1 = C_2 = C$ and $n = 1$. The divergence between two classes is:

$$J_{(1,2)} = \frac{(\mu_1 - \mu_2)^2}{\delta^2} \qquad (6)$$

In which μ_1 and μ_2 are means, and δ^2 is the variance.

Case 2. The two classes have the same means: $m_1 = m_2$ and $\delta = 0$. The divergence is:

$$J_{(1,2)} = \frac{1}{2} \mathrm{tr}\left[(C_1 - C_2)(C_2^{-1} - C_1^{-1}) \right]$$

$$= \frac{1}{2} \mathrm{tr}\left[(C_1 C_2^{-1}) \right] + \mathrm{tr}\left[(C_2 C^{-1}) \right] \qquad (7)$$

For single wheat kernel color classification, the largest distance between reflectance spectra of red wheats and white wheats must relate to the greatest dissimilarity of red wheats and white wheats. Therefore, the wavelength region with the largest distance between the reflectance spectra of red wheats and white wheats tends to cover the most important features which carry large discriminatory information and can be used as input to a neural network.

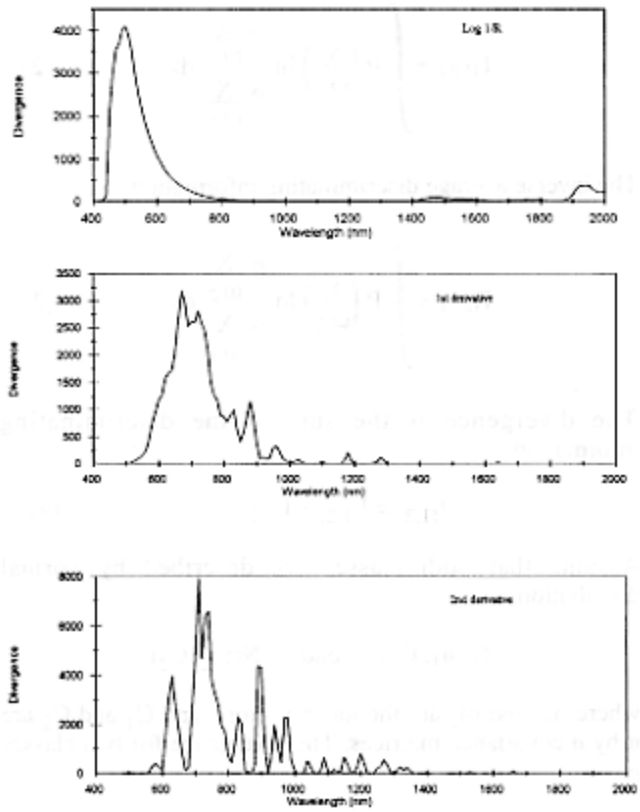

Figure 2–Divergence of single wheat kernel color class as a function of wavelength (nm).

Figure 2 shows the divergence curve between red wheats and white wheats with Log(1/R), the first derivative, and the second derivatives for sample sets without durum. The results indicate that the largest distance between two sample sets for Log (1/R) is between 450 nm and 850 nm. Therefore, the wavelength region from 450 nm to 850 nm could be used for neural network analysis. A complete discussion of divergence feature selection method is included in Tou and Gonzalez (1974).

Principal Components Analysis

Principal component analysis (PCA) is a mathematical procedure for resolving sets of data into orthogonal components whose linear combinations approximate the original data. In spectroscopy, the spectra data are usually correlated and the number of underlying components is smaller than the number of variables or number of spectra. The PCA method approximates a spectral vector with a linear combination of a set of orthogonal vectors, or factors as:

$$Y_i \cong a_{i1}C_1 + a_{i2}C_2 + a_{i3}C_3 + \ldots a_{ik}C_k \qquad (8)$$

where Y_i is a spectral vector, C_k is the k*th* factor (component), and a_{ik} is the k*th* coefficient associated with the i*th* spectrum. Coefficients a_{i1} to a_{ik} are called the scores of the ith spectral vector. In this way, the spectra in a wavelength space can be transformed into a vector space with k dimensions. Then the scores of each spectral vector can be used to develop neural network models.

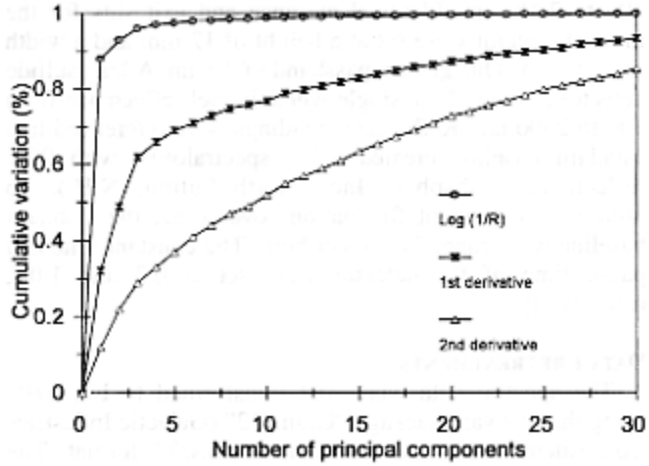

Figure 3–The cumulative variations as a function of the number of the principal components.

As successive components are calculated, the first few principal components account for the maximum possible amount of residual variance in the set of data. The number of principal components used for neural network analysis is based on the cumulative variations accounted for. Figure 3 shows the cumulative variation as a function of the number of principal components. For Log (1/R), the first 10 principal components account for about 99% of the total variation. Therefore, the first 10 principal components could be used for neural network analysis. For the first derivative, 30 principal components were needed to account for about 91% of the total variation. This means that when the first derivative is used for neural networks, more principal components are needed. For the second derivative, each principal component accounts for a small percent of the total variation. Thus, even more principal components are needed when using the second derivative. The PRINCOMP procedure (SAS Institute, Inc., 1990) was used to calculate principal components.

Neural Networks

The software package "NeuralWorks Professional II/Plus" (NeuralWare, Inc., Pittsburgh, Pa.) was used for data analysis based on back-propagation networks. Back-propagation uses a learning process to minimize the global error of the system by modifying node weights. The weight increment or decrement is achieved by using the gradient descent rule. The network is trained by initially selecting the weights at random and then presenting all training data repeatedly. Weights are adjusted after every trial using external information specifying the correct result until weights coverage and error are reduced to acceptable values. A complete discussion of back-propagation network theory is given by Hecht-Nielsen (1989).

The network inputs were selected by the divergence feature selection method and principal component analysis method as discussed earlier. Table 2 shows the experimental designs for neural network classification with the divergence feature selection method. The training sample sets and the testing sample sets used for principal components extraction are the same as the sample sets used for neural network analysis based on the divergence feature selection method. For the PCA method, each sample's

Table 2. Experimental design for neural network classification of single wheat kernel color class using visible and NIR reflectance spectra with divergence feature selection method

Classification Models	Wavelength Range (nm)	Data Points
Sample set without durum*		
Log(1/R)	450-850	201
1st derivative	550-950	201
2nd derivative	600-1000	201
Sample set with durum*		
Log(1/R)	450-800	176
1st derivative	550-950	201
2nd derivative	600-1000	201

* For the sample set without durum, the number of kernels in the calibration set are: red = 225 and white = 150; the number of kernels in the testing set are: red = 350 and white = 250. For the sample set with durum, the number of kernels in the calibration set are: red = 225 and white = 225; the number of kernels in the testing set are: red = 350 and white = 325.

scores associated with the first 10, 15, and 20 principal components from decomposition of the 500 to 1700 nm region were calculated. These PCA scores were used as input nodes to develop neural network models. For each classification experiment, two types of neural networks (with and without a hidden layer) were tested for each data pretreatment. The network with the highest testing accuracy was recorded as the best model.

Results and Discussion

The neural network classification models were based on two sample sets (sample set with or without durum) and three different data pretreatments [Log (1/R) and the first and second derivatives]. Dowell (1994) and Bochereau et al. (1992) showed no benefit of using more than one hidden layer. Thus, only the difference between one hidden layer and no hidden layer networks were compared.

Neural Network Input Reduction Using the Divergence Feature Selection Method

The effect of hidden layer nodes and a hidden layer on single wheat kernel color classification was determined using Log (1/R) values from 375 kernels for the training set and 600 kernels for the testing set with 40,000 learning events over the wavelength region of 450 to 850 nm (table 3). The results indicated that the network with one hidden layer did not improve the classification accuracy even when a different number of hidden layer nodes, which agrees with results published by Chen et al. (1995) and Song et al. (1995), was used. Better performance occurred when there was no hidden layer. Therefore, the following discussion and results refer to the network models without a hidden layer.

The major parameters that affect the classification and the training speed of the neural networks are the number of learning cycles (events), learning rate (η), and momentum (α). Table 3 shows the effect of learning rate and momentum on the classification accuracy. The results show that when the momentum was kept constant (M = 0.4), the highest learning rate resulted in the lowest classification accuracy. Learning rate determines how much of the error to propagate back into the preceding nodes, and it affects the speed of convergence of the network. A lower learning rate (< 0.4) may be slow because of small weight changes, and may need more learning cycles to achieve the required error level. A high learning rate (> 0.6) may not produce convergence. Therefore, the learning rate of 0.6 and momentum of 0.4 were used for network training. The learning cycles used during network training were based on the network with the highest testing accuracy.

The testing results for two sample sets with three different data pretreatments using the wavelength regions listed in table 2 are summarized in table 4. For both sample sets, the average classification accuracy ranged from 96.4% to 98.8%, and the average number of learning events for each record was from 65 to 100. For both sample sets, the highest classification accuracy was obtained from Log (1/R), followed by first derivative. The highest accuracies were 98.3% for the sample set without durum and 98.8% for the sample set with durum.

Table 3. Effect of hidden layer nodes and one hidden layer on color classification when 375 kernels were used as a training set and 600 kernels served as testing set. The network was trained for 40,000 learning events over the wavelength region of 450 to 850 nm

	Misclassified Kernels				
		One Hidden Layer with Different Number of Nodes			
Learning Rate / Momentum	No Hidden Layer	5	10	20	40
	Learning rate (Momentum = 0.4)				
0.1	13	27	27	27	27
0.4	13	26	26	25	28
0.6	13	25	24	24	28
0.9	23	30	28	29	31
	Momentum (Learning rate = 0.6)				
0.1	13	28	27	29	31
0.4	13	25	24	24	28
0.6	13	27	26	28	30
0.9	14	26	27	27	27

Table 4. Testing results for a neural network when the divergence feature selection was used

Calibration Model / Pretreatment	n*	Classification Accuracy (%) Red†	White†	Avg.	Calibration Cycle / Spectrum (Total Cycle)
Sample set without durum					
Log(1/R)	600(10)‡	97.4	99.2	98.3	65 (24375)
1st derivative	600(15)	97.4	97.6	97.5	96 (36000)
2nd derivative	600(20)	96.6	96.8	96.7	100 (37500)
Sample set with durum					
Log(1/R)	675(8)	98.3	99.2	98.8	75 (33750)
1st derivative	675(19)	96.0	95.5	97.2	85 (38250)
2nd derivative	675(24)	95.1	97.8	96.4	80 (36000)

* Number of kernels.

† Red = hard red spring + hard red winter + soft red winter; White = hard white wheat + soft white wheat for sample set without durum; White = hard white wheat + soft white wheat + durum for sample set with durum.

‡ Values in parentheses = the number of kernels misclassified.

Neural Network Input Reduction Using Principal Components Analysis

Two types of neural network models were tested. One neural network model had 10, 15, and 20 input nodes without a hidden layer. The other model had one hidden layer with 3, 7, and 7 hidden nodes, respectively, for the neural network model with 10, 15, and 20 input nodes.

The testing results are shown in table 5. The classification accuracies were similar between the neural network with a hidden layer and without a hidden layer. The highest classification accuracy for the sample set without durum was 98% obtained from Log (1/R) with 20 input nodes and using the network without a hidden layer. For the sample set with durum, the highest classification accuracy was 96.3% obtained from Log (1/R) with 10 input nodes and using the network without a hidden layer. Among the data pretreatments, Log (1/R) yielded the highest classification accuracy and the second derivative yielded the lowest classification accuracy (less than 80%) for the two types of neural network models (with or without a hidden layer).

Figure 4 shows the first three extracted principal components from the training set with Log (1/R). The first principal component, which accounted for the most variance of the data set, was probably due to the color variation between red wheat and white wheat. Because the first principal component in the NIR region of 800 to 1700 nm is almost flat, it can be considered that the first component is not related to any major constituents of wheat such as protein, moisture, oil, starch, or cellulose. The second principal component, which had large scores around 590 nm, may be caused by the visible color of the kernel. The third principal component, which had large scores around 590 nm and 1570 nm, was dominated by color characteristics (590 nm) and protein and hardness

Table 5. Testing results of a neural network using principal components as inputs

	Classification Accuracy (%) for No. of Input Nodes		
Pretreatment	10	15	20
Sample set without durum*			
No hidden layer			
Log 1/R	95.2	95.2	98.0
1st derivative	88.8	87.3	87.0
2nd derivative	69.7	68.2	69.2
With hidden layer			
Log 1/R	94.8	95.3	95.0
1st derivative	88.7	87.5	86.3
2nd derivative	68.6	69.2	69.8
Sample set with durum*			
No hidden layer			
Log 1/R	96.3	94.7	95.3
1st derivative	91.6	90.7	89.2
2nd derivative	75.7	78.5	78.8
With hidden layer			
Log 1/R	95.8	95.4	95.8
1st derivative	90.1	90.5	89.6
2nd derivative	76.1	78.8	79.7

* For the sample set without durum, the number of kernels in the testing set are: red = 350 and white = 250. For the sample set with durum, the number of kernels in the calibration set are: red = 225 and white = 225; the number of kernels in the testing set are: red = 350 and white = 325.

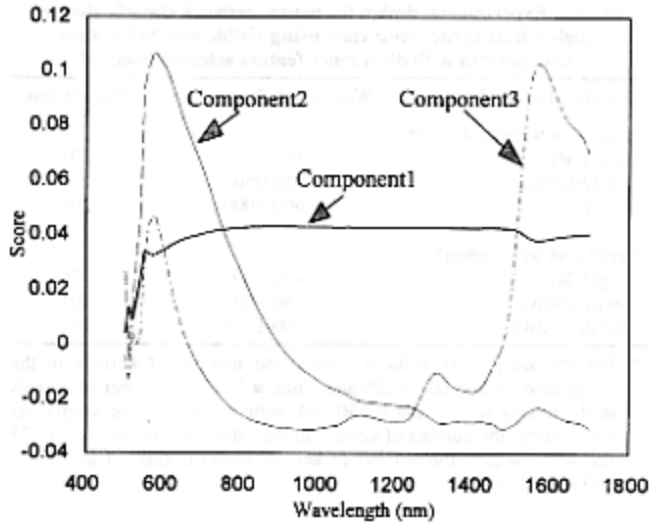

Figure 4–First three principal components (factors) extracted from the training set with Log (1/R).

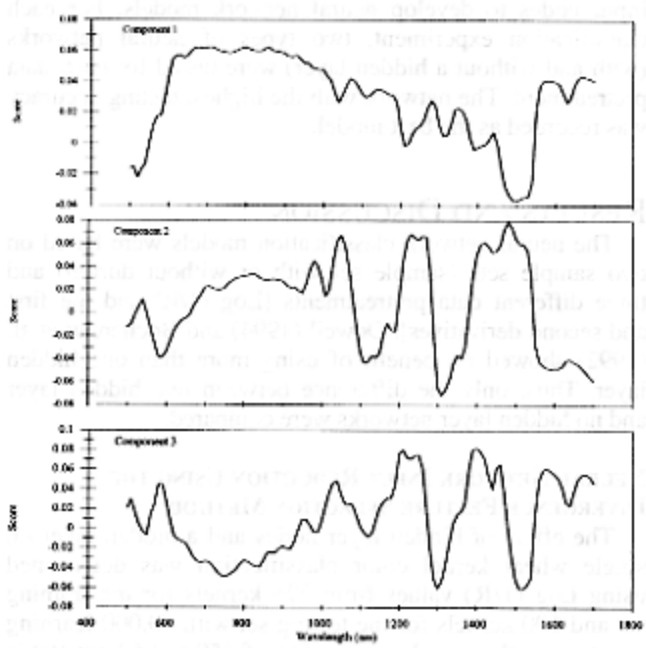

Figure 5–First three principal components (factors) extracted from the training set with the first derivative.

(1570 nm) (Murray and Williams, 1990; Burns and Ciurczak, 1992). However, the third principal component accounted for a small amount of variation of the data set when compared to the first two principal components.

Figure 5 shows the first three principal components extracted from the training set with the first derivative. The first principal component, which had large scores from 600 to 1000 nm, was dominated by variation due to color class. The divergence feature selection method showed that the largest distance between red and white wheats for the first derivative is from 550 to 950 nm (table 2). The score peaks and valleys (1132, 1246, 1300, 1438 nm and others) may be related to wheat color class or interaction of the major functional group such as protein, starch, water, oil, and cellulose. The other peaks and valleys in the first,

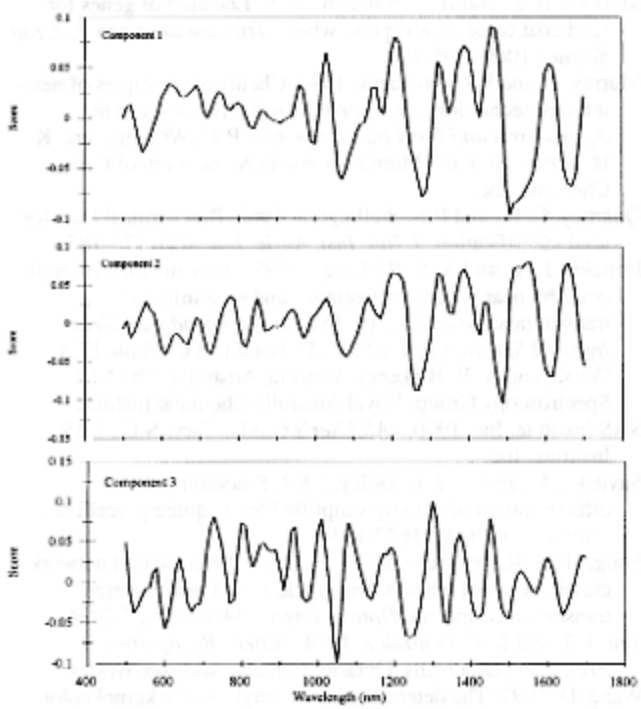

Figure 6–First three principal components (factors) extracted from the training set with the second derivative.

second, and third principal components may relate to protein, oil, starch, hardness, and interaction of the major constituents of wheat.

Figure 6 shows the first three principal components extracted from the training set with the second derivative. Score peaks and valleys around 560, 650, 750, 788, 814, and 840 nm are related to wheat color class (Wang, 1997). But these score peaks and valleys are weaker than the other score peaks and valleys in the longer wavelength region. The strong score peaks and valleys that occurred in the first three principal components are mostly related to protein, oil, and kernel hardness. For example, the score peaks and valleys around 1200, 1390, and 1480 nm are strongly related to protein content in the wheat (Delwiche, 1996; Delwiche and Massie, 1996). The score peaks and valleys around 960, 1060, 1330, 1390, 1480, and 1680 nm are strongly related to the kernel hardness of wheat, and 1390 nm is also related to oil content of wheat (Delwiche, 1993, 1996; Delwiche and Massie, 1996; Murray and Williams, 1990; Burns and Ciurczak, 1992). These strong score peaks and valleys related to protein, oil, and kernel hardness likely reduced the classification accuracy of red and white wheats when PCA was used.

Table 6 shows the effect of kernel hardness on the percentage of misclassified kernels when PCA was used to reduce the neural network input dimension. For red wheats, soft wheat had a high percentage of misclassified kernels. For white wheats, hard wheat had a high percentage of misclassified kernels. These results indicate that kernel hardness affects wheat color classification.

Table 6. Effect of kernel hardness on percentage of misclassified kernels when principal component analysis was used for neural network input reduction (500-1700 nm)

	Misclassified Kernels (%)	
Data Pretreatment	Hard Wheat	Soft Wheat
Log (1/R)		
Red wheat	4.8	10.0
White wheat	12.0	0.0
1st derivative		
Red wheat	1.6	21.0
White wheat	16.8	4.0
2nd derivative		
Red wheat	23.0	33.0
White wheat	38.4	32.0

Conclusions

1. Divergence feature selection and principal components analysis methods are effective for neural networks input data dimension reduction.
2. The accuracies for single wheat kernel color classification ranged from 96.4% to 98.8% for different data pretreatments when the divergency feature selection method was used for data dimension reduction.
3. The accuracies for single wheat kernel color classification ranged from 68.2% to 98% for different data pretreatments and the number of principal components used when the principal components analysis method was used to for data dimension reduction.
4. For both data reduction methods, Log (1/R) yielded the highest classification accuracy and the second derivative yielded the lowest classification accuracy.

References

Baker, R. J. 1981. Inheritance of seed coat color in eight spring wheat cultivars. *Canadian J. Plant Sci.* 61(7): 719-721.

Bason, M. L., S. Zounis, J. A. Ronalds, and C. W. Wrigley. 1995. Segregating red and white wheat visually and with a tristimulus color meter. *J. Agric. Res.* 46(1): 89-98.

Bjørsvik, H.-R., and H. Martens. 1992. Data analysis: Calibration of NIR instruments by PLS regression. In *Handbook of Near-infrared Analysis*, eds. D. A. Burns, and E. W. Ciurczak. New York, N.Y.: Marcel Dekker Inc.

Bochereau, L., P. Bourgine, and B. Palagos. 1992. A method for prediction by combining data analysis and neural networks: Application to prediction of apple quality using near infrared spectra. *J. Agric. Engng. Res.* 52(1): 207-216.

Burns, D. A., and E. W. Ciurczak. 1992. Principles of near-infrared spectroscopy. In *Handbook of Near-infrared Analysis*, eds. D. A. Burns, and E. W. Ciurczak. New York, N.Y.: Marcel Dekker.

Chen, C. Y., S. K. Skarsaune, and C. A. Watson. 1972. Relation of kernel color to wheat class and grade. *Cereal Sci. Today* 17(11): 340-343.

Chen, Y. R., S. R. Delwiche, and W. R. Hruschka. 1995. Classification of hard red wheat by feedforward backpropagation neural networks. *Cereal Chem.* 72(3): 317-319.

Chmelar, F., and K. Mostovoj. 1938. On the application of some old and on the introduction of new methods for testing genuineness of variety in the laboratory. *Proc. Int. Seed Test. Assoc.* 10(1): 68-73.

Delwiche, S. R. 1993. Measurement of single-kernel wheat hardness using near-infrared transmittance. *Transactions of the ASAE* 36(5): 1431-1437.

Delwiche, S. R., and H. K. Norris. 1993. Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. *Cereal Chem.* 70(1): 29-35.

Delwiche, S. R. 1996. Single kernel protein content in wheat by near-infrared reflectance. ASAE Paper No. 96-3031. St. Joseph, Mich.: ASAE.

Delwiche, S. R., and D. R. Massie. 1996. Classification of wheat by visible and near-infrared reflectance from single kernels. *Cereal Chem.* 73(3): 399-405.

DePauw, R. M., and T. N. McCaig. 1988. Utilization of sodium hydroxide to assess kernel color and inheritance in eleven spring wheat varieties. *Canadian J. Plant Sci.* 68(4): 323-329.

Dowell, F. E. 1994. Neural network parameter effects on object classification and wavelength selection. *Optical Engng.* 33(8): 2728-2732.

Dowell, F. E. 1997. Effect of NaOH on visible wavelength spectra of single wheat kernels and color classification efficiency. *Cereal Chem.* 74(5): 617-620.

Flintham, J. 1993. Grain color and sprout-resistance in wheat. In *Proc. 6th Int. Symp. Pre-Harvest Sprouting Cereals*. St. Paul, Minn.: American Association Cereal Chemists.

Gunst, R. F., and R. L. Mason. 1979. Some consideration in the evaluation of alternate prediction equations. *Technometrics* 21(1): 55-63.

Hawk, A. L., H. H. Kaufmann, and C. A. Watson. 1970. Reflectance characteristics of various grains. *Cereal Sci. Today* 15(11): 381-384.

Hecht-Neilsen, R. 1989. *Neural Computing*. New York, N.Y.: Addison-Wesley Publishing Co.

Hruschka, W. R. 1990. Data analysis: Wavelength selection methods. In *Near-infrared Technology in the Agricultural and Food Industries*, eds. P. C. Williams, and K. H. Norris. St. Paul, Minn.: American Association of Cereal Chemists, Inc.

Kimber, G. 1971. The inheritance of red grain colour in wheat. *Pflanzenzuecht* 66(1): 151-157.

Massie, D. R., and K. H. Norris. 1965. Spectral reflectance and transmittance properties of grain in the visible and near infrared. *Transactions of the ASAE* 8(4): 598-600.

McCaig, T. N., J. G. McLeod, J. M. Clarke, and R. M. DePauw. 1992. Measurement of durum pigment with a near-infrared instrument operating in the visible range. *Cereal Chem.* 69(6): 671-672.

McCaig, T. N., R. M. DePauw, and P. C. Williams. 1993. Assessing seed-coat color in a wheat breeding program with a NIR/VIS instrument. *Canadian J. Plant. Sci.* 73(4): 175-182.

Metzger, R. J., and B. A. Sibaugh. 1970. Location of genes for seed coat color in hexaploid wheat, *Triticum aestivum* L. *Crop Science* 10(5): 495-496.

Murray, I., and P. C. Williams. 1990. Chemical principles of near-infrared technology. In *Near-infrared Technology in the Agricultural and Food Industries*, eds. P. C. Williams, and K. H. Norris. St. Paul, Minn.: American Association of Cereal Chemists, Inc.

Quartley, C. E., and P. S. Wellington. 1962. Biochemical tests for seed identification. *J. Nat. Inst. Agric. Bot.* 9(2): 179-185.

Ronalds, J. A., and A. B. Blakeney. 1995. Determination of grain color by near infrared reflectance and near infrared transmittance spectroscopy. In *Leaping Ahead with Near Infrared Spectroscopy*, eds. G. D. Batten, P. C. Flinn, L. A. Welsh, and A. B. Blakeney. Victoria, Australia: The NIR Spectroscopy Group, Royal Australia Chemical Institute.

SAS Institute, Inc. 1990. *SAS User's Guide*. Cary, N.C.: SAS Institute, Inc.

Savitsky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least square procedures. *Anal. Chem.* 36(8): 1627-1639.

Song, H., S. R. Delwiche, and Y. R. Chen. 1995. Neural network classification of wheat using single kernel near-infrared transmittance spectra. *Optical Engng.* 34(10): 2927-2934.

Tou, J. T, and R. C. Gonzalez. 1974. *Pattern Recognition Principles*. Don Mills, Ontario, Canada: Addison-Wesley.

Wang, D. 1997. The determination of single wheat kernel color class using visible and near-infrared reflectance. Ph.D. diss. College Station, Tex.: Texas A&M University.

William, R. H. 1990. Data analysis: Wavelength selection methods. In *Near-Infrared Technology in the Agricultural and Food Industries*, eds. P. C. Williams, and K. H. Norris. St. Paul, Minn.: American Association of Cereal Chemists, Inc.